Bulletin Technique n° 16
Institut d'Intelligence Artificielle
Université Paris VIII

L'EFFET "PHRASES CACHÉES"

(méthodes d'analyse stylistique comportant la prise en considération de structures syntaxiques cachées ou virtuelles)

Yves LECERF

GENERALITES : PROBLEMATIQUE DE L'"EFFET PHRASES CACHEES"

Sous le nom générique d'"effet", dans les sciences physiques, il est habituel de désigner des constructions déductives combinant à la fois des observations expérimentales et des hypothèses théoriques, concernant des phénomènes qu'avant leur observation on aurait pu juger assez inattendus, voire même insolites.

Utilisée ici maintenant dans un domaine expérimental assez différent, celui de la linguistique, cette dénomination d'effet qui va être appliquée à ce que nous désignerons ici comme "l'effet phrases cachées" ne nous a pas semblé cependant mal adaptée à son objet ; car les phénomènes mis en cause par "l'effet phrases cachées" concernent une catégorie de chaînes syntaxiques que les linguistes n'ont pas coutume de considérer avec attention : à savoir des chaînes non syntagmatiques, découpées de façon assez quelconque dans des phrases ; et c'est un fait dans une certaine mesure insolite que de telles chaînes puissent avoir des propriétés non aléatoirement quelconques.

On se souviendra en effet que l'effort patient de la linguistique générale puis structurale durant nombre de décennies a tendu précisement au contraire à démontrer que les langues naturelles ne sont pas isotropes au fractionnement ; à démontrer qu'il n'est ni pertinent ni légitime de procéder à des découpages quelconques de la chaîne syntaxique ; à démontrer qu'au contraire un fractionnement légitime du discours ne peut s'effectuer que selon certains segments privilégiés dits "syntagmes" et absolument pas selon un fractionnement quelconque en des chaînes qui inclueraient la fin d'un sytagme et le début d'un autre. Pourtant, avec ce que nous appellerons "l'effet phrases cachées" c'est précisement ce genre de chaînes aberrantes que nous allons considérer, les chaînes non syntagmatiques, en vue de décrire leurs propriétés.

Encore, parmi ces chaînes, ferons nous une place toute particulière à une catégorie à priori aberrante : la catégorie des chaînes susceptibles à elles seules de constituer une phrase complète si on les extrait de leur contexte, avec une signification autonome, allant souvent à contresens du document dont on les a extraites ; à ces chaînes, nous donnerons le nom de "phrases cachées".

Les occurences ne manquent pas de ce genre de phrases cachées dans des textes de toutes origines, et l'on voit, et l'on sait à quel point des auteurs, hommes politiques, orateurs,ont pu déplorer souvent la pratique journalistique de la "citation tronquée". Pour illustrer bien toutefois le mécanisme de la phrase cachée, nous mentionnerons quelques exemples :

la phrase : "Jean épouse Jeanne et Christine épouse Christian"	contient la phrase : "Jean épouse Jeanne et Christine"
la phrase : "les voleurs fuyant devant ces chiens policiers totalisaient vingt agressions à main armée"	contient la phrase : "ces chiens policiers totalisaient vingt agressions à main armée"
la phrase : "la robe que j'avais achetée à Adrienne avait complètement rétréci sous la pluie"	contient la phrase : "Adrienne avait complètement rétréci sous la pluie"
la phrase : "le chat de la fermière croque une souris"	contient la phrase : "la fermière croque une souris"

Et il apparaît clairement que des exemples de ce genre pourraient être sans difficulté multipliés à l'infini.

Quelle importance, quelle pertinence, attribuer à cet ensemble de chaînes non syntagmatiques que sont les "phrases cachées"?

C'est toute la problématique de "l'effet phrases cachées" que nous nous proposons d'exposer ici : un point de vue selon lequel ces phrases constitueraient une sorte de face cachée du langage, dont la prise en considération pourrait éclairer la compréhension de phénomènes linguistiques et notamment stylistiques qui sont habituellement considérés comme directement pertinents. Retraçons rapidement l'histoire de cette problématique.

Il n'existe pas à notre connaissance d'étude de psychanalyse du langage qui ait directement pris pour thème les "phrases cachées" ; l'occasion ne s'en est sans doute pas présentée ; mais on peut tenir cependant pour vraisemblable que ces phrases auraient, dans un tel éclairage, été considérées comme porteuses possibles de signification subconsciente.

Toujours est-il que les premiers concours de circonstances qui ont conduit à faire, des "phrases cachées" et des "segments cachés" y afférents, l'objet d'une véritable démarche de réflexion logique et scientifique, datent vraisemblablement des environs des années 1960, avec l'avènement de l'informatique, et la mise en oeuvre d'une première série importante de recherches sur l'analyse algorithmique des langues naturelles dans l'espoir d'applications ultérieures à la traduction automatique des langues.

Parmi les nombreuses difficultés rencontrées dans le déroulement des travaux d'analyse syntaxique automatique des langues, figuraient des phénomènes parasites dus à l'existence des "phrases cachées" et des segments analogues, qui introduisaient souvent des contresens dans les analyses, et conduisaient dans tous les cas à une consommation de temps d'ordinateur.

Par le terme générique de "phrases cachées", nous voulons en effet désigner ici un phénomène d'ensemble, qui porte aussi sur de simples segments pseudosyntaxiques cachés : chaînes nominales cachées, chaînes adjectivales cachées, etc. par exemple :

le segment de phrase :
"le poudrier d'une dame de l'assistance dont le dos est émaillé ..." (I,3)
(d'après Bally)

contient les segments :
"une dame de l'assistance dont le dos est émaillé ..." (I,4)
"l'assistance dont le dos est émaillé..." (I,5)

A propos de ce dernier exemple, on peut se demander si la syntaxe interdit réellement que le dos d'une dame soit émaillé et, sachant que des ordinateurs sont seulement capables d'appliquer strictement des consignes syntaxiques, mais non de tenir compte de la signification des phrases, l'on peut se faire une idée de l'ampleur des difficultés introduites en analyse automatique par l'existence du phénomène de "phrases cachées". Malgré l'importance des efforts de recherche déployés dans le monde depuis plus de dix ans, ces difficultés sont très loin d'avoir pu être surmontées, du point de vue de l'exactitude des réponses données par les machines.

Les phénomènes parasites de "phrases cachées" n'affectaient, dans ces expériences d'analyse syntaxique automatique, pas seulement le résultat du calcul mais aussi, comme déjà dit, sa durée.

Il arrivait souvent, par exemple que tout paraisse fonctionner normalement: les machines ayant reçu des phrases à analyser, avaient abondamment fait clignoter leurs lumières sur les consoles de contrôle, elles avaient fait tourner leurs bandes et leurs disques avec le doux ronronnement d'un bon fonctionnement, elles avaient mené à bien leurs travaux et rendu des réponses exactes. Mais des reconstitutions de l'emploi du temps des unités de calcul, effectuées après coup, montraient que ces organes avaient longuement raisonné sur des "phrases cachées" du type : "la fermière croque une souris" cité plus haut pour conclure d'ailleurs qu'il fallait les rejeter;ceci d'autant plus longuement que pour certaines phrases, les possibilités de fausses pistes étaient vraiment très nombreuses,affectant par exemple simultanément plusieurs groupes de mots : fausses chaînes nominales, fausses chaînes verbales, etc. par exemple :

la phrase :
"le chat de la fermière de mon père croque tous les matins une souris" (I,6)

contient les phrases :
"la fermière de mon père croque tous les matins une souris" (I,7)
"mon père croque tous les matins une souris" (I,8)
"la fermière de mon père croque tous les matins" (I,9)
"mon père croque tous les matins" (I,10)
"le chat de la fermière de mon père croque tous les matins" (I,11)

Le fait qu'une phrase aussi courte et aussi simple puisse contenir déjà rien moins que cinq"phrases cachées" toutes bien construites donne une idée de

l'ampleur que peut atteindre le phénomène et du nombre des possibilités combinatoires examinées par les ordinateurs : ainsi, dans les résultats, rien n'apparaissait des maudites "phrases cachées" mais tout se passait comme si les machines avaient rêvé longtemps à leur sujet.

Pour en terminer avec cet exposé historique, on peut rappeler que les difficultés ainsi rencontrées du fait des "phrases cachées" et de leurs interférences avec les analyses automatiques de phrases normales correspondent à des expériences s'étendant des années 1960 environ jusqu'aujourd'hui puisqu'elles durent encore. Vue sous l'angle spécifique de ces difficultés, et sans approfondissement des concepts en cause, la problématique du phénomène "phrases cachées" se résumerait en quelques mots : un effet parasite, résultant d'un ensemble de petits faits antérieurement considérés comme totalement insignifiants.

Les raisons que nous pouvons éventuellement mettre en avant aujourd'hui, de renverser complètement cette problématique, et d'affirmer que l'effet"phrases cachées" constitue une face cachée du langage, et que, dans la mise en oeuvre du langage, cet "effet phrases cachées" peut impliquer un rôle fonctionnel précis, ces raisons sont à rechercher à la fois sur le plan de la théorie et de l'observation.

Sur le plan de la théorie : s'il s'agit d'éliminer "l'effet phrases cachées" en tant que phénomène parasite, il semble raisonnable de proposer des modèles de calculs d'identification des phrases qui intègrent exclusivement ce phénomène : l'objet de l'algorithme devient alors de calculer à la fois les structures de phrases et les structures de "phrases cachées".

Or nous proposons justement un certain type de modèle algébrique (Algèbre de semi anneau sur un magma) qui présente naturellement cette propriété d'intégrer les calculs de "phrases cachées" (sous forme d'un ensemble de calculs de signification à résultante nulle) au calcul normal. En outre, ce type de représentation algébrique permet d'intégrer naturellement au calcul un certain nombre d'autres structures, ce qui constitue a posteriori une raison supplémentaire de lui attribuer une valeur explicative meilleure que celle de représentations n'intégrant pas les "phrases cachées".

Sur le plan de l'observation : les faits observés à l'occasion d'expériences d'analyse automatique du langage par ordinateur, les difficultés rencontrées du fait des phrases cachées dans le déroulement de ces travaux, tout cela reste insuffisant pour inspirer un renversement de la problématique du phénomène "phrases cachées" : car seuls y sont en cause des ordinateurs, et des méthodes de calcul dont rien ne prouve qu'elles s'apparentent suffisamment à celles de l'intelligence humaine.

Pour affirmer que "l'effet phrases cachées" est une sorte de face cachée du langage, il faut des bases expérimentales d'observation supplémentaires. Il faut montrer que les "phrases cachées" peuvent être impliquées dans des aspects fonctionnels précis du langage, par exemple dans ses aspects stylistiques.

Que la stylistique des "phrases cachées" varie d'un auteur à l'autre, nul n'en doutera à priori et c'est d'ailleurs un fait que l'on peut vérifier. Que la prise en considération de l'ensemble des "phrases cachées" et segments cachés, au lieu de la prise en considéraiton du texte sous sa face directe, puisse constituer une sorte de vision radiographique singulière, éventuellement assez pénétrante en tant qu'instrument d'analyse stylistique, nul n'en doutera a priori, non plus, et c'est d'ailleurs aussi un fait que l'on peut vérifier, sans qu'il soit question

d'exagérer la portée de cette méthode qui restera un simple instrument d'analyse parmi beaucoup d'autres. Tout cela donnerait en tout cas déjà une base d'argumentation pour proposer de renverser la problématique du phénomène "phrases cachées".

Mais que l'on puisse en venir à certaines observations laissant supposer que les "phrases cachées" joueraient quelquefois un rôle fonctionnel direct et directement explicable dans certains effets de style, c'est alors là que des raisons deviendraient vraiment importantes de voir dans les"phrases cachées" une seconde face cachée du langage. Or nous pensons que des apparences très fortes, des présomptions très fortes, existent en faveur d'un rôle direct des "phrases cachées" dans une certaine forme de style poétique : d'une part, les "phrases cachées" et "segments cachés" apparaissent en quantité anormalement élevée, et d'autre part ces"segments cachés" présentent des caractéristiques esthétiques, comme si l'on avait tiré parti de leur existence pour cacher en eux des densités poétiques supplémentaires.

A cet ensemble d'observations et de considérations théoriques que nous venons d'énumérer, il nous semble au total justifié de donner le nom proposé en début de ce chapitre, de problématique de "l'effet phrases cachées".

DES POEMES CACHES DANS DES "PHRASES CACHEES" DANS DES POEMES

Il existe au moins deux raisons évidentes de présumer que les textes poétiques en général peuvent constituer un terrain d'étude privilégié pour le phénomène des "phrases cachées".

L'une des raisons tient aux usages sémantiques de la littérature poétique, qui font à l'insolite une assez bonne place, surtout à l'insolite dans des présentations à moitié hermétiques du fait d'un certain caractère caché ; et les quelques exemples de mécanismes de "phrases cachées" cités plus haut ont montré que ces phrases pouvaient avoir des significations quelquefois étranges.

La seconde de ces raisons tient, au contraire, non plus à la sémantique mais à la forme des textes de littérature poétique : ces textes sont précisément entre tous, les seuls à mettre en évidence à l'intérieur de leurs propres phrases une segmentation supplémentaire complètement anormale, qui est celle de la versification. Cette désinvolture formelle des"phrases cachées" à l'égard de la syntaxe des phrases normales qui les hébergent, syntaxe que les"phrases cachées" prennent délibérément en biais ou à contre sens, nous la retrouvons dans la désinvolture de versification de maints poèmes :

"Douces colonnes, aux
Chapeaux garnis de jour" (Valéry)

"Parmi les marronniers, parmi les
Lilas blancs, les lilas violets" (Moréas)

Quant à l'organisation méthodique des études que l'on peut envisager de conduire sur le rôle des"phrases cachées"dans la littérature poétique c'est très naturellement qu'elle trouve son point d'appui et son point de départ dans une recherche exhaustive des "phrases cachées" dans les textes poétiques.

Pour dresser ainsi la liste de ces "phrases cachées", il pourrait sembler normal et naturel d'utiliser des ordinateurs puisqu'aussi bien leur aptitude particulière à détecter de tels segments est justement à l'origine de notre démarche ; et l'on pourrait espérer que leur puissance de calcul permette de traiter rapidement une large collection de textes.

Cependant, le parti que nous devrons adopter consistera précisément, au contraire à nous contenter de techniques artisanales, appliquées à un certain nombre de textes très courts choisis comme échantillons : les "phrases cachées" seront recherchées si l'on peut dire "à la main", sans moyen mécanique ou électronique auxiliaire, mais en reproduisant artisanalement et dans la mesure du nécessaire, les opérations qu'aurait effectuées un ordinateur. Cette renonciation à l'emploi de moyens mécaniques s'explique par plusieurs raisons, dont la principale tient à l'état des "programmes" actuellement existants dans le monde pour l'analyse syntaxique automatique de textes en langue française : il s'agit seulement de programmes expérimentaux, dotés chacun d'un vocabulaire et d'une syntaxe de dimensions très restreintes ; ces programmes sont utiles pour observer expérimentalement à petite échelle certains phénomènes, mais demeurent totalement impropres à un quelconque traitement de masse de textes français ; il est donc hors de question de mettre en oeuvre une expérience de recherche massive de "phrases cachées" par ordinateur.

Ajoutons que la perspective de réaliser, pour l'avenir, une expérience massive de ce genre demeure d'autant plus éloignée que le corpus à traiter est celui des textes poétiques, c'est-à-dire le corpus de textes dont la syntaxe présente les pires difficultés imaginables du point de vue d'un dispositif d'analyse grammaticale automatique.

Du reste, le fait que la recherche "artisanale", manuelle, des "phrases cachées" demeure aujourd'hui et pour longtemps la seule modalité pratiquement envisageable n'entraîne pas nécessairement que cette procédure soit dépourvue d'avantages et d'efficacité. Le choix d'un nombre suffisant d'échantillons, certes assez courts, peut nous donner certainement une première approximation des phénomènes étudiés, un peu comme un "sondage" peut donner, à partir d'échantillons, une approximation d'une situation statistique. Pour ce qui concerne le dépouillement de chaque texte, on s'efforcera d'imiter la démarche des ordinateurs, en observant que leur efficacité à découvrir des "phrases cachées" tient tout simplement au caractère exhaustif des découpages de segments usuellement commandés par les programmes d'analyse : il y aura donc lieu de procéder tout simplement à un découpage systématique, combinatoire et exhaustif des textes, suivi d'un examen également systématique de chaque segment. Certains segments révèleront la présence de "phrases cachées", et l'on en dressera l'inventaire.

Un avantage du caractère artisanal du procédé est de permettre de consigner dans ce même inventaire diverses autres particularités notées pour chacun des segments rencontrés, particularités par exemple de versification qui relèvent de critères non syntaxiques et n'auraient pas été détectables par des programmes d'analyse automatique ; ou bien détection de certaines catégories très particulières de "phrases cachées" non mécaniquement inventoriables en tant que n'ayant pas la structure de phrases bien constituées.

Dans la catégorie des segments acceptables au titre de "phrases cachées", il semble en effet raisonnable de faire une place à certains segments comparables dans leur structure à ces phrases courtes et lapidaires que sont les "titres de livres" comme par exemple "Pour qui sonne le glas", "Vipère au poing","Sous le soleil de Satan". De tels segments sont indiscutablement placés"entre un point et un point"dans les textes imprimés, comme des phrases normales. Les textes poétiques peuvent en

contenir, comme on verra plus loin, une assez importante proportion ; par exemple, trois en un seul vers d'Eluard :

"... Pointe à frayeur, fruit en guenilles, amour rapace ..."

Nous verrons, à l'occasion d'exemples donnés plus loin, que la signification de certains de ces segments après leur découpage va parfois complètement à l'encontre de la signification du texte dont ils sont extraits ; ce qui est la preuve de leur qualité de "phrases cachées" autonomes, de la même façon que "la fermière croque une souris" constituait un contresens par rapport à la signification principale "le chat de la fermière croque une souris" dans un exemple déjà cité.

Ces opérations de recensement de l'ensemble des "phrases cachées" et "segments cachés" une fois menées à bien, il reste à en tirer parti dans des observations stylistiques.

Du point de vue qualitatif, il faudra apprécier l'ensemble des segments ainsi sélectionnés, et cela implique sans doute une certaine part d'arbitraire. Toutefois une mention particulière doit être donnée des cas éventuels où certains segments-vers en forme classique (alexandrins), etc. - pourraient être détectés dans les "phrases cachées" car l'identification de structures de ce genre ne peut être considérée comme relativement objective. Mention particulière également doit être donnée des "segments titre de livre" dont il vient d'être parlé plus haut, car eux aussi se prêtent à une appréciation relativement extérieure au domaine de la poésie, donc relativement objective : la question de savoir si un segment donné pourrait faire ou non un bon titre de livre met en jeu une intuition relativement mécanique, elle appelle facilement une réponse par oui ou par non, avec de bonnes concordances d'opinion entre plusieurs sujets interrogés. C'est un test que nous utiliserons largement dans la suite de cette étude.

Du point de vue quantitatif, il sera bon de comparer des textes poétiques et des textes en prose ; c'est ce que nous ferons pour mettre en évidence la différence de densité en "segments cachés" qui sépare ces deux modes d'expression stylistique.

DES MICRO-POEMES EN FORME DE NOMS DE POEMES OU NOMS DE LIVRES OU NOMS DE TABLEAUX CACHES DANS DES POEMES

Il a déjà été fait allusion plus haut à cette catégorie très particulière de phrases que sont les titre de poèmes, les titres de livres, les noms de tableaux, et l'on a fait observer que de tels segments pouvaient être à bon droit considérés comme des phrases, certes très courtes. Il semble justifié de les prendre en compte dans des inventaires de "phrases cachées".

C'est avant tout un style, riche en oppositions et violent dans ses images, qui caractérise ce genre de segments que sont les titres ; et il y a en effet certainement une gageure à essayer de donner un pouvoir d'évocation à des segments aussi courts. Pourtant les titres ont ce pouvoir : "Sang et lumière" ; "Une fille de Saragosse" ; "Tendre et violente Elisabeth" ; "Par qui le scandale" ; "Soudain l'été dernier" etc.

Si l'on omet dans une évaluation de leur longueur les mots incolores tels que "une" "de" "la" etc. on observe que les titres sont dans l'ensemble construits sur deux ou trois notions sémantiques principales ; par exemple trois pour "La

grande peur dans la montagne", pour "Baigneuse du clair ou sombre",pour "Blason des fleurs et des fruits", pour "Le quatrième poème visible" ; et par exemple deux pour "Mort en fraude", "Les anges noirs", "Les violons parfois" ; et même quelquefois une seule notion comme dans "La peau", "La peste", "Olivia", "Le mépris". En tous cas, les noms de livres sont des segments courts et doivent, pour des raisons mnémotechniques, être tels.

Si, inversant le problème, nous posons la question de savoir ce que pourrait être un genre poétique de poèmes très courts, soumis à la contrainte d'un maximum de trois notions, il nous semble qu'un tel genre aurait beaucoup de parenté avec la littérature des titres de livres. L'extrême brièveté du genre imposerait que l'on recherche là aussi des compensations soit par le moyen d'oppositions très contrastées, soit par le choix d'images violentes ou riches en évocation, soit encore par le recours à des découpages insolites. Et quant à l'idée même de poème très court, elle n'est pas si absurde, puisque la littérature japonaise connaît avec le Haiku un genre de poème à cinq ou six notions.

Nous pensons donc que des segments en forme de titres de poèmes, titres de livres, noms de tableaux, trouvés dans des poèmes ordinaires au hasard des textes, présenteraient le double intérêt d'être d'une part une certaine forme de"phrases cachées" et d'autre part une certaine forme de"micropoèmes cachés".

A leur sujet, nous énonçons une affirmation : la fréquence des"phrases cachées" en forme de titres de poèmes,titres de livres, noms de tableaux, rencontrés dans les textes poétiques dépasse nettement celle atteinte dans les autres textes.

Et sur ce même sujet, nous ajouterons que cette singularité est un des traits qui contribuent le plus à créer dans les textes poétiques la situation de "multiple fractionnabilité" dont la description fait l'objet de la présente étude. En effet, les noms de poèmes ou de livres ont une structure, on l'a vu, très concise et ramassée : "Croisade sans croix", "Pleure ô mon pays bien aimé", "Le baiser au lépreux". Par ailleurs, ils ne constituent pas nécessairement une chaîne d'apparence syntagmatique normale : "Les violons parfois", "Par qui le scandale" ; "Dans un mois, dans un an", "Soudain l'été dernier", etc.

La vérification d'une propriété de fractionnabilité selon des segments de ce genre laisse sans doute quelque marge d'appréciation à la subjectivité. Soit l'exemple d'un quatrain d'Eluard (A la flamme des fouets) :

> "Métal qui nuit, métal de jour, étoile au nid,
> Pointe à frayeur, fruit en guenilles, amour rapace,
> Porte couteau, souillure vaine, lampe inondée,
> Souhaits d'amour, fruits de dégoût, glaces prostituées ..."

Il nous semble quant à nous que tous les segments constitutifs du texte (segments intercalés entre virgules, tous sans exception) pourraient faire de remarquablement bons titres de romans ou poèmes. De la même façon une lecture du "Cantique des colonnes" de Valéry permet de relever des : "hymnes candides", "servantes sans genoux", "filles des nombres d'or", "un dieu couleur de miel" auxquels nous prêtons la même qualité ; et de même une lecture du poème "A la nue accablante tu" de Mallarmé nous donne des segments tels que : "de basalte et de laves", "échos esclaves", "trompe sans vertu", "sépulcral naufrage", "perdition haute", et ainsi de suite. Mais pour tous ces segments, nous ne sommes pas en mesure de citer de précédent défini de livre, poème, film, tableau, portant précisément ces noms.

Ce que l'on peut affirmer, relativement à de tels précédents, c'est que sans aucun doute des emprunts sont constamment faits au corpus des textes poétiques existants par des auteurs cherchant des titres pour des oeuvres nouvellement composées.

"Le sel de la terre"	(Gide)	origine : nouveau testament
"Par qui le scandale"	(Hougron)	origine : nouveau testament
"Bonjour tristesse"	(Sagan)	origine : Eluard

Sans doute serait-il souhaitable de disposer d'un inventaire exhaustif des emprunts de ce genre ou même des coïncidences involontaires ; mais de toute manière, l'ensemble des segments de poèmes est beaucoup plus vaste que celui des titres déjà attribués à d'autres oeuvres ; et un inventaire de ce genre ne suffirait donc même pas à apprécier le phénomène dans toute son ampleur; il faudrait de toute manière avoir recours à l'appréciation stylistique de segments comme titres possibles.

Puisqu'aussi bien le titre "Bonjour tristesse" vient d'être mentionné comme extrait d'un poème d'Eluard, considérons ce poème : il s'intitule "A peine défigurée" et son texte est le suivant :

"Adieu tristesse
Bonjour tristesse
Tu es inscrite dans les lignes du plafond
Tu es inscrite dans les yeux que j'aime
Tu n'es pas tout à fait la misère
Car les lèvres les plus pauvres te dénoncent
Par un sourire
Bonjour tristesse
Amour des corps aimables
Puissance de l'amour
Dont l'amabilité surgit
Comme un monstre sans corps
Tête désappointée
Tristesse beau visage"

En dehors du segment "Bonjour tristesse" effectivement choisi comme titre d'un roman, on observe que bon nombre de chaînes de mots ont le style des titres de livres : "Adieu tristesse", "les lignes du plafond", "les yeux que j'aime", "la misère", "les lèvres les plus pauvres", "les plus pauvres te dénoncent", "Par un sourire", "Amour des corps aimables", "Puissance de l'amour", "monstre sans corps", "Tête désappointée" ; et même aussi "tristesse bonjour", "tristesse tu es", "sourire bonjour", "des corps aimable puissance", bien que ces dernières chaînes soient un peu moins bonnes ; en outre, bon nombre des vers du poème semblent acceptables en qualité spécifique de titres de poèmes mais peut-être moins facilement de titres de livres : c'est le cas par exemple du segment "tristesse beau visage" dont le caractère irrationnel s'accomoderait mieux d'un poème que d'un roman.

Il est important de le faire observer, les significations évoquées par certains de ces segments se différencient très nettement de celles constituant la trame du poème ; par exemple "des corps aimables puissance" est probablement un solécisme et un contresens, car puissance se réfère à "puissance de l'amour" et non aux corps aimables ; par exemple encore "les plus pauvres te dénoncent" est un solécisme et évoque une image peut-être voisine du texte mais tout de même assez distincte ; par exemple encore "sourire bonjour" est un solécisme et sa signification n'a guère de rapport avec celle du passage dont on le tire, et ainsi de suite.

Une caractéristique de la"phrase cachée" consiste en effet précisément dans ceci qu'elle possède une autonomie propre de signification ; par exemple la signification de "la fermière croque une souris" s'écartait nettement de celle de la phrase d'origine "le chat de la fermière croque une souris". En retrouvant cette autonomie chez les segments titres de livres, nous obtenons une confirmation supplémentaire de leur aptitude à constituer de véritables "phrases cachées" et non plus seulement des fragments de phrases normales.

Ce phénomène d'autonomie de signification des segments découpés est observable dans bon nombre de poèmes d'Eluard. Par exemple dans le quatrain :

"Une ou plusieurs
L'azur couché sur l'orage
La neige sur les oiseaux
Les bruits de la peur dans les bois revêches".

Chacun des vers peut être considéré comme un titre de livre possible et même assez brillant : "une ou plusieurs", "l'azur couché sur l'orage", "la neige sur les oiseaux","les bruits de la peur dans les bois revêches" ; mais en outre d'autres découpages du texte suggèrent des titres à contresens : "sur l'orage la neige", "sur les oiseaux les bruits de la peur" dont l'existence confirme le statut d'autonomie de signification des segments : ce sont bien des "phrases cachées".

Toujours dans ce même quatrain, notons que l'énumération du sens et du contre-sens n'épuise pas complètement la liste des "segments titres de livres": des chaînes telles que "couché sur l'orage" ou "la peur dans les bois" ou "les bruits de la peur" etc. constituent des titres acceptables, de même encore que "l'azur couché", "les bois revêches", "sur l'orage" etc. Certes, elles ne forment pas des contre-sens caractérisés, mais on ne saurait exiger que toutes les "phrases cachées" aillent à contre-sens de la signification principale. Observons enfin que le total des segments notés pour ce simple quatrain devient tout de même considérable si l'on se réfère aux dimensions modestes du texte dont ils sont extraits : leurs longueurs cumulées dépassent largement le double de la longueur du texte.

Les traits observés seraient-ils particuliers aux poèmes d'Eluard? Il n'en est rien bien entendu. Prenons au hasard un poème de Baudelaire, par exemple la très célèbre "Invitation au voyage":

"Mon enfant, ma soeur
Songe à la douceur
D'aller là-bas vivre ensemble!
Aimer à loisir
Aimer et mourir
Au pays qui te ressemble!
Les soleils mouillés
... etc."

Ici encore le texte est constitué presque en totalité de segments susceptibles de servir de titres à des livres : "Enfant ma soeur", "songe", "douceur d'aller", "songe à la douceur", "là-bas vivre ensemble", "aimer à loisir", "ensemble aimer", "aimer et mourir", "au pays qui te ressemble", "mourir au pays", "à loisir aimer", "les soleils mouillés" etc. Et ici encore, l'autonomie de la signification est évidente pour des titres tels que "mourir au pays" lequel évoque immédiatement un souci très répandu de dormir son dernier sommeil nulle part ailleurs qu'au cimetière du village natal, donc le contraire de l'invitation au voyage, le contraire du sens principal. Et ici encore, la densité du texte en segments dans le style de titres de livres apparaît comme très riche.

Un inventaire que l'on voudrait vraiment exhaustif de tous les segments de textes susceptibles ainsi de servir de noms à d'autres oeuvres, poèmes, romans, tableaux conduirait parfois à retenir la presque totalité des segments, c'est à dire même

pour des phrases très courtes un nombre assez considérable de chaînes. Prenons l'exemple d'une phrase courte choisie au hasard dans le roman "Dedans" d'Hélène Cixous, au milieu de la première page du livre:

"Dans mon jardin d'enfer les mots sont mes fous".

On peut découper dans cette phrase plusieurs segments très typiques du style des titres de livres en général : "jardin d'enfer", titre qui suggère un récit fantastique à la Poe ; "les mots", titre réel d'un ouvrage de Sartre ; "d'enfer les mots", qui fait songer aux aphorismes d'Esope sur le langage ; "les mots sont", titre qui nous renvoie à la caverne de Platon et à la querelle des universaux. La diversité même de ces significations montre une fois de plus que les "segments titres de livres" sont de véritables phrases cachées, aussi distinctes du sens principal que notre "phrase cachée" type "la fermière croque une souris" citée plus haut pouvait l'être de sa phrase mère "le chat de la fermière croque une souris". Toujours dans la même phrase d'Hélène Cixous, d'autres titres peuvent encore être trouvés : "dans mon jardin", évocateur de joies bucoliques ; "les mots sont mes fous", hermétique. "Enfer" qui est presque un titre de Dante ; "Fous", pour citer les segments les plus évidents seulement. Un découpage exhaustif de la phrase conduit à découvrir encore quelques segments acceptables supplémentaires mais qui ne sont que des variantes des précédents ; par exemple : "jardin" et "mon jardin", distincts du segment "Dans mon jardin" déjà cité, avec d'ailleurs de nettes possibilités de divergence de signification ; par exemple "Mots" au lieu de "les mots" déjà cité ; par exemple "d'enfer" au lieu de "enfer" déjà cité , etc.

Au total, on parvient sans doute à découvrir quelques segments vraiment inacceptables, tels que "d'enfer les " ou "dans mon", segments dont un découpage combinatoire révèle l'existence et qui ne peuvent en aucun cas être rattachés au style des titres de livres ; mais il est remarquable d'observer que ces segments constituent plutôt l'exception. Le test de la recherche de phrases cachées par découpage de la phrase initiale d'Hélène Cixous se révèle donc extrêmement positif.

Mais cet exemple, tiré à la différence des précédents d'un texte en prose, pose le problème de la distinction entre prose poétique et prose tout court.

"L'EFFET PHRASES CACHEES" ET LA PROSE

C'est en l'opposant à"la prose" que l'on peut donner de la littérature poétique les définitions les plus simples, dans des formulations qui toutefois recèlent bien des dangers. C'est en tout cas dans "la prose", indiscutablement, que réside l'ultime élément de preuve, dont nous avons lieu de faire état pour démontrer le caractère singulier des phénomènes de "microphrases cachées"décelés dans de nombreux textes poétiques : cette singularité particulière, il faut montrer que les textes "en prose" ne l'ont pas.

C'est, il est vrai, à dessein que nous avons choisi comme dernier exemple de texte poétique, un texte en prose, le roman "Dedans" d'Hélène Cixous : il y a des textes en prose qui relèvent du statut de la littérature poétique, il y a des poèmes en prose,il y a de la prose poétique ; bref, il y a parfois des poèmes cachés dans de la prose.

Compte tenu de la complexité de cette situation, une discussion devra nécessairement suivre l'examen du ou des contre-exemples choisis dans "de la prose qui soit vraiment de la prose" ; mais nous suggérons qu'il soit procédé d'abord à cet examen des textes eux-mêmes, en commençant par une lettre adressée le 6 mars 1697 à M. Ménage, par un de nos plus admirables prosateurs français, Mme de Lafayette :

> "J'attends présentement Clélie avec toute l'impatience que j'ai accoutumé d'avoir pour elle ; mais cette impatience est encore augmentée par les louanges que vous lui donnez. Quand vous louez quelque chose pour être bien écrit, il faut s'en fier à vous ; car, pour l'ordinaire, vous ne louez guère que ce qui mérite d'être loué. Je pourrais tirer de là une conséquence qui me serait assez avantageuse ; car vous m'avez tant louée en français et en italien que je serais en droit de croire que j'aurais quelque sorte de mérite. Mais ma vanité ne me mène pas si loin et j'attribue les louanges que vous m'avez données à l'amitié que vous avez pour moi, sans y chercher d'autre cause"

Il est selon notre sentiment impossible de découper le moindre titre de livre un peu incisif ou percutant de l'ensemble de ces lignes. A la rigueur, les quatre premiers mots "J'attends présentement Clélie", pourraient-ils rappeler l'ennui voulu d'un "En attendant Godot", mais de bien loin ; sans doute aussi, certains mots isolés pourraient-ils constituer des titres d'ouvrages, mais d'une formulation bien terne : "amitié" "l'amitié" "impatience" "mérite" "louanges" ; à la rigueur aussi quelques groupes de mots tels que "l'impatience que j'ai accoutumé d'avoir" ou bien "quelque sorte de mérite", mis en valeur par un découpage insolite et par l'insolite de découper une banalité ; mais on voit bien que ces quelques possibilités de titres n'ont de justification que par antiphrase, autrement dit parce qu'il ne s'agit pas de titres normaux ; et encore sont-ils bien insignifiants. Le décompte total des "titres de livres" possibles est donc bien proche, pour un découpage dans ce texte, d'un état néant.

Pour s'en convaincre mieux encore, l'on peut soumettre une des phrases, par exemple la troisième, à un fractionnement exhaustif :

> "Quand vous/vous louez/quand vous louez/louez quelque chose/vous louez quelque chose/louez quelque chose pour/quelque chose pour être/quelque chose pour être bien/pour être/pour être bien/pour être bien écrit/être bien écrit/bien écrit,il/écrit,il faut/écrit il faut s'en/il faut s'en fier/faut s'en fier/faut s'en fier à/s'en fier à vous/fier à vous/".
> "Quand/vous/louez/quelque/chose/pour/être/bien/écrit/il/faut/s'en/fier/à/vous/."

Il n'y a pas de segment acceptable pour un titre dans cet ensemble sauf peut-être, à la rigueur, certains mots isolés : par exemple "être". C'est peu.

Le "contre exemple" ayant ainsi apporté confirmation de son caractère différent des "exemples" cités avant lui, exemples de poèmes dont on pouvait extraire bon nombre de segments en forme de noms de poèmes, noms de livres, noms de tableaux, il nous reste à discuter de cette distinction qui justement existe entre "prose poétique" et "prose tout court".

Une théorie séduisante consisterait à poser en principe que la propriété faisant l'objet de notre étude peut justement constituer un critère de cette distinction : mériterait le nom de "prose poétique" toute prose dans laquelle il serait possible de discerner de nombreux fragments ayant l'éclat d'un "minipoème" en forme de nom de poème, et ceci même de préférence, en sorte que le découpage soit continu.

Quant à l'incidence de "l'effet phrases cachées" dans les textes en prose, il ne faudrait pas se hâter de conclure qu'elle est nécessairement nulle : on peut penser dans la mesure où il ne s'agit pas de prose poétique, ni de poèmes en prose, que cette incidence prend d'autres formes, qui varient selon les auteurs.

Sans doute avons-nous pu constater que le style de Madame de Lafayette présente une assez faible densité de "phrases cachées" allant à contre-découpage et à contre-sens du texte, comme en "phrases cachées" du type "titres de livres". Son cas peut être rapproché de celui d'un autre prosateur féminin, auteur célèbre en littérature enfantine, la Comtesse de Ségur ; et l'on conçoit du reste assez bien qu'il en soit ainsi pour cette dernière ; l'existence de "phrases cachées" à signification ésotérique ou à contresens ne serait certainement pas de nature à favoriser la facilité de lecture pour un jeune public. Pour ce qui concerne les propriétés de fractionnabilité et segmentabilité de la prose de la Comtesse de Ségur, on aura lieu d'observer qu'elle reste curieusement élevée, mais cette segmentabilité conduit à des chaînes dont les significations confirment chaque fois celles des phrases dont elles sont extraites,et dont les structures confirment celles normalement assimilables aux phrases ; en outre chaque chaîne exprime le plus souvent une affirmation qui pourrait être considérée comme vraie isolément, par référence au sens général du texte.

Exemple de texte :

> "Après donc m'avoir tenu les propos les plus humiliants, auxquels je ne répondais qu'en me léchant les lèvres et en lui tournant le dos, elle prit son bâton et se mit à me battre si cruellement que je finis par perdre patience et que je lui lançai trois ruades, dont la première lui cassa le nez et deux dents, la seconde lui brisa le poignet, et la troisième l'attrapa à l'estomac et la jeta par terre"
> (Comtesse de Ségur : "Les mémoires d'un âne")

Segments remarquables :
"je ne répondais qu'en me léchant les lèvres et en lui tournant le dos"
"elle prit un bâton et se mit à me battre"
"je finis par perdre patience"
"je lui lançai trois ruades"
"la première lui cassa le nez et deux dents"
"la seconde lui brisa le poignet"
"la troisième l'attaqua à l'estomac et la jeta par terre"

Ces segments remarquables sont grosso modo tous des syntagmes du texte, mais non pas des phrases ni même des propositions indépendantes coordonnées : et cependant la situation est bien telle que si des parties de texte échappent à l'attention des jeunes lecteurs, les parties remarquables qui restent comprises sont interprétables comme des phrases et leur sens confirme celui du texte.

UN MODELE ALGEBRIQUE POSSIBLE POUR LA REPRESENTATION DE "L'EFFET PHRASES CACHEES"

Parmi les formalismes que l'on peut proposer pour représenter à la fois une phrase et ses "phrases cachées", l'un des plus naturels est celui de la fonction Π, "produit contracté dans un magmoïde". On trouvera une description détaillée de ce formalisme dans l'article "Méthodes de calcul pour une analyse informatique des structures syntaxiques cachées de textes poétiques"("Effet phrases cachées").

Indiquons brièvement ici ce que nous entendons par magmoïde : c'est une algèbre de semi anneau A définie sur un magma E, avec opération extérieure de multiplication par des pseudo scalaires appartenant à un semi anneau α défini sur un second magma ε.

Indiquons d'autre part ce que nous entendons par produit contracté Π. La définition récursive de cette fonction est la suivante :

$$\Pi(a_1 a_2 \ldots a_n) = \sum_{i=1}^{i=n-1} \Pi(\overset{i}{\rightarrow} a_1 a_2 \ldots a_n) \; \Pi(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$$

$$\Pi(a_j) = a_j \quad \text{pour tout } a_j \in E$$

formule dans laquelle les produits et les sommes doivent s'interpréter en termes d'opérations du semi anneau E ; dans laquelle $(\overset{i}{\rightarrow} a_1 a_2 a_3 \ldots a_n)$ désigne la chaîne $(a_1 a_2 \ldots a_n)$ limitée à ses i premiers termes;et dans laquelle $(a_1 a_2 \ldots a_n \overset{n-i}{\leftarrow})$ désigne la chaîne $(a_1 a_2 \ldots a_n)$ limitée à ses n-i derniers termes.

Expliquons pourquoi ce formalisme est proposé.

Une des manières les plus simples que l'on puisse trouver de représenter une famille de structures de phrases ou de fragments de phrases, consiste à recourir à des structures d'arbres articulées autour des éléments (monèmes, mots) qui constituent ces phrases ou fragments ; et la structure algébrique de magma permet de représenter un arbre par un terme monome produit non associatif ; une structure de semi anneau (algébrique) A sur un magma E permettra donc de représenter des familles d'arbres dans un formalisme polynomial.

Pour arriver à faire des tris de sous-catégories et de sous familles d'arbres, en sorte à parvenir à pondérer les énumérations d'arbres par des coefficients susceptibles de s'annuler, il est nécessaire d'introduire une opération de multiplication extérieure, par exemple par un scalaire ; mais on voit bien que la multiplication par des pseudo scalaires relevant eux aussi d'une algèbre de semi anneau α sur magma ε ne présente pas plus de difficulté formelle, car le calcul des pseudo scalaires prend appui sur les formalismes de toutes façons déjà nécessaires pour les éléments du magma E déjà existant ; et le fait de disposer de pseudo scalaires (à table de multiplication arbitraire) donne des possibilités de tris bien supérieures à celles des scalaires simples pour sélectionner des sous ensembles d'arbres dans E.

Dans le cas particulièrement simple où E est un magma libre et ε un magma fini, la structure de magmoïde se rapproche beaucoup d'une description de grammaire context-free, les éléments de E figurant des arbres et leurs coefficients étant égaux à la catégorie syntaxique (non terminale) du syntagme correspondant à chaque arbre.

Quant au produit contracté Π, ce serait une fonction égale à la somme de toutes les structures d'arbres possible, s'il n'était le rôle des coefficients pseudo-scalaires nuls qui font que Π se réduit à un sous ensemble particulier ; en l'absence de coefficients nuls, Π regroupe vraiment toutes les structures : c'est une fonction universelle de compilation.

Le fait de choisir une représentation de phrase par des arbres peut sembler de nature à introduire une certaine limitation mais il faut bien noter que les pseudo scalaires extérieurs sont utilisables en nombre quelconque (sur des magmas auxiliaires $\varepsilon_1 \varepsilon_2 \varepsilon_3 \ldots \varepsilon_p$ etc.) et que ces pseudo scalaires sont potentiellement aussi des arbres;autrement dit, l'élément de base de cette structure n'est pas l'arbre simple mais le produit cartésien de plusieurs arbres.

On en trouvera dans le document déjà cité en référence une description beaucoup plus détaillée, ainsi qu'une étude des propriétés principales de ce système formel, notamment sur l'aptitude à représenter des contraintes complexes faisant intervenir de façon combinée et imbriquée des contraintes relevant à la fois de plusieurs grammaires context-free différentes ; notamment aussi son aptitude à représenter des langages complètement non context-free, tels que le langage écho x-x.

On trouvera dans un autre article, intitulé "Méthodes de calcul pour une évaluation de l'effet versification cachée dans la versification" des détails sur la possibilité de représenter, par ces moyens, des contraintes introduisant des relations complexes entre des structures de versification et des structures syntaxiques. Le champ d'application de ce type de calculs s'étend notamment à la caractérisation de structures versifiées éventuellement incluses dans des phrases cachées quelconques, elles sont données par la fonction Π, sous réserve d'ajuster convenablement les coefficients.

DES PERCEPTIONS SUBCONSCIENTES DE FRAGMENTS DE PHRASES CACHEES DANS LA PERCEPTION CONSCIENTE DES PHRASES ; DES CHAINES FAUSSEMENT SYNTAXIQUES CACHEES DANS DES RAISONNEMENTS SYNTAXIQUES CORRECTS

S'il n'est pas invraisemblable que la qualité poétique des parties constituantes d'un texte, prises selon de multiples fractionnements entrecroisés, contribue dans un sens favorable à la qualité d'ensemble du texte dans son entier, il reste que l'on est en droit de s'interroger et de rester perplexe sur ce qui concerne la possibilité effective pour l'auditeur ou le lecteur du poème de percevoir à la fois tant de "beautés" dans tant de fragments.

Des phrases cachées dans des phrases ; des poèmes cachés dans des poèmes ; des "micro-poèmes" en forme de noms de poèmes, noms de livres, noms de tableaux, noms de romans, cachés dans des poèmes ; des poèmes cachés dans de la prose ; des versifications insolites cachées dans des phrases cachées ; que perçoit-on vraiment de tout cela?

Certes, les fragments d'un texte, et tous ses fragments, sont dans la situation de se trouver énoncés lorsque le texte est prononcé (et dans la situation d'être parcourus par l'oeil lorsque le texte est lu). Un auditeur frappé d'une amnésie telle qu'il ne pourrait mémoriser que les quatre ou cinq derniers mots prononcés devant lui, serait en cette qualité d'amnésique tout particulièrement apte à percevoir les fragments d'un poème en tant qu'unités distinctes. Mais pour un auditeur normal, le fait que les fragments soient énoncés ne suffit pas à démontrer que ces fragments sont perçus en tant que tels.

Quant au caractère à exiger d'une telle perception, il semble raisonnable de se contenter éventuellement d'une perception subconsciente. S'agissant de sentiment esthétique, une perception consciente n'est en effet pas à priori indispensable.

Il ne serait donc pas indifférent de pouvoir apprécier la fréquence usuelle de perception subconsciente, l'importance usuelle de perception subconsciente des éléments poétiques cachés dans un poème ou cachés dans de la prose. S'il pouvait être établi que la perception subconsciente de ces "éléments cachés" est usuelle fréquente, importante, alors l'on serait sans doute conduit à formuler l'hypothèse que ces éléments poétiques cachés jouent directement un rôle fonctionnel déterminé dans le "charme poétique". S'il pouvait au contraire être établi que la perception subconsciente des éléments poétiques cachés est toujours nulle, alors, l'on serait conduit à formuler l'hypothèse que ces éléments cachés n'ont pas de rôle fonctionnel direct dans le "charme poétique".

Nous ne disposons guère d'éléments factuels susceptibles de permettre l'appréciation de cette éventuelle perception subconsciente ; du moins, guère d'éléments pour cette perception subconsciente par des gens ; car au contraire, nous disposons d'éléments factuels très nombreux relatifs à des machines, relatifs à des situations qu'à tort ou à raison nous estimons comparables, et où la perception est le fait d'une machine et non pas d'une personne.

Nous l'avons dit plus haut, l'actuelle problématique de "l'effet phrases cachées" et du rôle des "phrases cachées" dans la littérature poétique tire sa source d'observations effectuées à propos d'expériences d'analyse automatique de textes par ordinateur. Dans ces expériences portant sur des phrases ordinaires, on décelait des phénomènes parasites de "phrases cachées" : par exemple, la phrase déjà citée : "le chat de la fermière croque une souris" se révélait contenir la phrase cachée : "la fermière croque une souris", laquelle phrase cachée ne manquait pas d'être perçue et analysée par les programmes pour être ensuite rejetée comme non compatible avec les mots immédiatement précédents. Le travail ainsi effectué pour l'analyse des "phrases cachées" pouvait être cause de pertes notables en temps machine, ceci d'autant plus que le phénomène commençait au niveau de simples segments faussement syntaxiques, puis s'étendait de là à des combinaisons de ces segments, lesquelles combinaisons pouvaient se révéler assez nombreuses.

Ainsi, l'un des plus difficiles problèmes, parmi tous ceux auxquels s'est trouvée confrontée la recherche en analyse automatique de textes, a consisté et consiste encore à essayer en quelque sorte d'empêcher la "perception" des "phrases cachées" par les machines ; au moins à limiter les conséquences de cette perception en temps perdu à raisonner sur ces pseudo-phrases.

Il est particulièrement instructif de constater que, malgré plus de dix années d'efforts déployés dans les équipes de recherche de diverses universités du monde, il n'a nulle part été possible d'éliminer les phénomènes parasites liés à la prise en considération par les machines des "phrases cachées". Il s'agissait et s'agit encore pourtant uniquement d'un problème de programmation et de syntaxe à la fois : les sous programmes chargés de regrouper les mots en syntagmes et en phrases sont généralement très bien séparés de ceux remplissant d'autres fonctions, on les appelle programmes de "parsing", et c'est de leur organisation que relève l'élimination des phrases ou syntagmes parasites. De nombreuses améliorations ont été proposées pour ces techniques de parsing, d'autres les sont encore ; certaines des modifications que l'on peut apporter à ces programmes tendent par exemple à accélérer les raisonnements sur les "phrases cachées" ou bien à rendre ces raisonnements partiellement implicites au lieu d'explicites. Mais toutes les tentatives que l'on a pu faire pour sauter purement et simplement l'étape du raisonnement sur les "phrases cachées" se sont traduites immanquablement par l'omission d'une partie des réponses correctes : mélangées aux "phrases cachées", une partie des phrases correctes, des solutions exactes, étaient éliminées sans examen ; et les analyses effectuées par les machines, si elles avaient l'avantage de se dérouler plus rapidement, le compensaient par l'inconvénient de donner moins souvent de bonnes réponses.

La prise en considération en tant que telles des pseudo phrases cachées constitue-t-elle une étape nécessaire au raisonnement syntaxique correct? Autrement dit, l'énonciation, au moins implicite, de l'erreur pour justement l'écarter constitue-t-elle une étape du raisonnement correct en analyse syntaxique automatique? L'expérience semble montrer qu'il en est ainsi. Pour des cas particuliers, le raisonnement permet même de l'affirmer. Pour certaines phrases en effet, empêcher l'unité de calcul de formuler certaines hypothèses fausses puis de les éliminer, c'est adopter un dispositif obligeant cette même unité de calcul à se tromper dans l'interprétation d'autres phrases pour lesquelles ces mêmes hypothèses se trouveraient au contraire

justement fondées. En voici des exemples :

le matin de la pluie serait bienvenu (VI,1)
la venue de la pluie serait bienvenue (VI,2)
le matin de la pluie dont je vous ai parlé ... (VI,3)
le matin de la pluie précéda l'émeute (VI,4)
le matin de la pluie avait précédé le matin de l'émeute (VI,5)

Dans les phrases (VI,1 à 5) ci-dessus, la chaîne "de la pluie" est tantôt complément de nom (VI,2,3 et 5) tantôt chaîne nominale sujet (VI,1 et 4), avec du reste une marge résiduelle non négligeable d'ambiguité. La chaîne "le matin" est tantôt nominale sujet (3,5) tantôt chaîne complément de temps (1,4). Quant à la question de savoir si la chaîne "le matin de la pluie" a une existence syntagmatique, sa réponse est tantôt oui (3,5) tantôt non (1,4).

L'analyse des phrases (VI,1 à 5) ne peut se faire rigoureusement qu'en essayant chaque fois pour "le matin de la pluie" les deux possibilités. La réussite dans une hypothèse ne dispense pas d'examiner l'autre, comme le montre la phrase tout à fait ambiguë :

"le matin de l'orage serait bienvenu" (VI,10).

L'histoire des techniques de parsing est riche d'exemples de phrases choisies par hasard comme phrases type et dont les auteurs découvraient après coup, avec la plus grande surprise, qu'elles comportaient une ambiguité (exemples : "le pilote ferme la porte", "they are flying planes" etc...).

Ainsi, la réponse correcte à un problème d'analyse syntaxique d'une phrase en langue naturelle inclut un double contenu : premièrement, la liste des interprétations syntaxiquement acceptables ; secondement, la démonstration qu'il n'y a pas d'autre interprétation syntaxiquement acceptable. Cette dernière démonstration peut souvent résulter pour certaines de ses parties de déductions implicites ; mais l'examen explicite de certaines chaînes virtuellement syntagmatiques (telle : "le matin de la pluie") en vue de les rejeter est un examen inévitable : l'énoncé de l'erreur pour la réfuter apparaît comme une étape nécessaire du raisonnement correct.

C'est là une des différences qui séparent le problème théorique de l'analyse des phrases du problème de la génération des phrases. Dans un processus génératif, par exemple pour la génération d'un langage de Chomsky, on ne considère que des chaînes syntagmatiques correctes :

S → NP + VP
NP → Art + N
VP → V + NP ... etc

Dans un processus d'analyse au contraire, on doit considérer toutes ces chaînes syntagmatiques correctes, plus un certain nombre de chaînes virtuellement syntagmatiques pour les éliminer. Ainsi, l'énoncé de l'erreur constitue une étape nécessaire du raisonnement correct en analyse.

Des raisonnements syntaxiques que nous venons de faire ci-dessus, et qui sont confirmés par l'expérience de nombreuses tentatives d'analyse de textes par ordinateurs, il convient de reporter maintenant les conclusions au problème dont nous étions partis, le problème de la perception subconsciente éventuelle de segments de textes poétiques : et nous voyons qu'il en résulte des présomptions importantes en faveur d'une réponse positive.

Certes le comportement de machines programmées ne peut en général pas être considéré comme un précédent indicatif de ce que serait le comportement d'êtres humains accomplissant des tâches comparables ; mais les contraintes que les problèmes de

machines ont fait découvrir ici sont des contraintes purement syntaxiques et logiques apparemment inhérentes à la nature du langage humain : de telles contraintes s'appliquent aux personnes aussi bien qu'aux machines.

Il semble donc probable que les "phrases cachées", phrases pseudo-syntaxiques et segments pseudo-syntaxiques cachés dans les phrases constituent une sorte d'écho intérieur du discours dans le discours, un écho qui serait perçu de façon inconsciente au cours du processus même de reconnaissance du langage. Et l'on ne saurait exclure qu'il puisse en résulter pour certains textes un effet esthétique.